

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 10:07:12 ; Search time 5166.57 Seconds
(without alignments)
-404.924 Million cell updates/sec

Title: US-09-276-268-2

Perfect score: 689

Sequence: 1 gtcgcctgaggtccgcgcg.....tgatccctcagcctctactg 689

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapert 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_bal.*

2: gb_ba2.*

3: gb_em.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pr1.*

10: gb_pr2.*

11: gb_pr3.*

12: gb_ro.*

13: gb_sts.*

14: gb_sy.*

15: gb_un.*

16: gb_vi.*

17: em_fun.*

18: em_hum1.*

19: em_hum2.*

20: em_in.*

21: em_om.*

22: em_or.*

23: em_ov.*

24: em_pat.*

25: em_ph.*

26: em_pl.*

27: em_ro.*

28: em_sts.*

29: em_sy.*

30: em_un.*

31: em_v1.*

32: gb_htg1.*

33: gb_htg2.*

34: gb_in1.*

35: gb_in2.*

36: em_bal.*

37: em_ba2.*

38: em_hum3.*

39: em_hum4.*

40: gb_pr4.*

41: gb_htg3.*

42: gb_htg4.*

43: gb_htg5.*

44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	327	47.5	2617	9	HSSGP1N15	Z50022 H.sapiens m
2	158	22.9	106687	43	AC015892	AC015892 Mus muscu
3	150.2	21.8	2230	10	HUMCTGREP	L48964 Homo sapien
4	118.8	17.2	178773	44	AC011782	AC011782 Homo sapi
5	118.8	17.2	333303	9	HSS171	AJ011930 Homo sapi
6	95.8	13.9	106687	43	AC015892	AC015892 Mus muscu
7	59.2	8.6	7218	5	I66494	I66494 Sequence 14
8	47.4	6.9	193387	11	AF064858	AF064858 Homo sapi
9	45.2	6.6	37655	40	AC007792	AC007792 Homo sapi
10	44	6.4	181098	11	AC004216	AC004216 Homo sapi
11	43.8	6.4	190837	11	HS90L6	Z97353 Human DNA s
12	43.6	6.3	122640	41	AC009401	AC009401 Homo sapi
13	43.4	6.3	281632	41	AC011407	AC011407 Homo sapi
14	43.4	6.3	81868	41	AC009689	AC009689 Homo sapi
15	43.4	6.3	105607	43	AC013410	AC013410 Homo sapi
16	43.2	6.3	2969	12	MUSLPL10	M60847 Mouse liopo
17	43.2	6.3	93145	43	AC012414	AC012414 Homo sapi
18	43.2	6.3	108064	41	AC010404	AC010404 Homo sapi
19	43	6.2	213581	42	AC009764	AC009764 Homo sapi
20	42.8	6.2	645	12	MMU239082	AJ239082 Mus muscu
21	42.6	6.2	160783	12	AC009287	AC009287 Mus muscu
22	42.4	6.2	379	13	HUMOT953	L18481 Human chrom
23	42.4	6.2	60332	43	AC012538	AC012538 Homo sapi
24	42.4	6.2	121978	43	AC016033	AC016033 Homo sapi
25	42.2	6.1	34612	44	AC016289	AC016289 Homo sapi
26	42.2	6.1	189906	33	HSJ738A13	AL109801 Homo sapi
27	41.8	6.1	4163	9	HSARPNGNA	Z24680 H.sapiens g
28	41.8	6.1	148340	42	AC011061	AC011061 Homo sapi
29	41.6	6.0	35000	44	AC016528	AC016528 Leishmani
30	41.6	6.0	168468	40	AC005510	AC005510 Homo sapi
31	41.6	6.0	175897	42	AC010723	G09243 human STS C
32	41.4	6.0	317	13	G09243	AC010723 Homo sapi
33	41.4	6.0	142075	40	AC005519	AC005519 Homo sapi
34	41.2	6.0	4238	12	RNCNG41	AJ000496 Rattus no
35	41.2	6.0	161987	40	AC006070	AC006070 Homo sapi
36	41	6.0	5166	12	AF122014	AF122014 Mus muscu
37	41	6.0	43900	11	AC005265	AC005265 Homo sapi
38	41	6.0	67841	44	AC016096	AC016096 Homo sapi
39	41	6.0	118195	40	AC007899	AC007899 Homo sapi
40	41	6.0	122383	11	HSJ944N18	AL049738 Human DNA
41	41	6.0	151490	45	AC011883	AC011883 Homo sapi
42	40.8	5.9	71342	43	AC015833	AC015833 Homo sapi
43	40.6	5.9	55520	11	HSJ488M4	AL078598 Human DNA
44	40.6	5.9	77516	10	AC002378	AC002378 Human PAC
45	40.6	5.9	129968	10	HS249F5	AL035691 Human DNA

ALIGNMENTS

RESULT 1

HSSGP1N15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

HSSGP1N15

H.sapiens mRNA for surface glycoprotein.

Z50022

Z50022.1

GI:1107702

2617 bp

mRNA

PRI

30-APR-1998

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 10:59:29 ; Search time 3493.2 Seconds
(without alignments)
669.053 Million cell updates/sec

Title: US-09-276-268-3
Perfect score: 619
Sequence: 1 ggcaccagggaagccctgcc.....tgtgcccaagggccactta 619

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:**
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	420.8	68.0	588	39	C86502
c 2	190.4	30.8	576	30	AA206019
					C86502 C86502 Mous
					AA206019 zq54d06.s

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 10:42:11 ; Search time 5166.57 Seconds
(without alignments)
-843.347 Million cell updates/sec

Title: US-09-276-268-6

Perfect score: 1435

Sequence: 1 catggcgccgtctgtag.....acacagctaaagcaacactg 1435

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pli.*
8: gb_pl2.*
9: gb_pri.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	582.6	40.6	2340	40	AF062534	AF062534 Homo sapi
2	429.8	30.0	113457	11	AC003686	AC003686 Homo sapi
3	142	9.9	339	13	G37764	G37764 SHGC-63501
C 4	57	4.0	7218	5	I66494	I66494 Sequence 14
5	46	3.2	159515	12	AC002327	AC002327 Mus muscu
C 6	45.8	3.2	3281	34	LMA243459	AJ243459 Leishmani
7	45	3.1	96106	45	AC017015	AC017015 Homo sapi
C 8	43.6	3.0	150079	41	AC009447	AC009447 Homo sapi
9	42.4	3.0	2685	10	HUMHBA3	J00184 Human alpha
10	42.4	3.0	43058	9	HSGG1	284721 Human DNA s
C 11	42.2	2.9	742	40	AF000672	AF000672 Homo sapi
C 12	42.2	2.9	2266	9	HUMELK1A	M25769 Homo sapien
C 13	42.2	2.9	29358	34	DMBH6115	AL035245 Drosophil
14	42.2	2.9	77656	45	AC017522	AC017522 Drosophil
C 15	41.6	2.9	25306	2	AF058302	AF058302 Streptomy
C 16	41.6	2.9	56414	1	MTV002	AL008967 Mycobacte
17	41.6	2.9	149977	10	CNS01DSV	AL122020 Human chr
C 18	41.6	2.9	174707	45	AC017082	AC017082 Homo sapi
C 19	41.6	2.9	232426	33	HS1086L22	AL109910 Homo sapi
20	41.4	2.9	202776	41	AC007871	AC007871 Homo sapi
C 21	41	2.9	42833	32	HS367G8	297634 Homo sapien
22	41	2.9	100516	32	HSS171M_3	Continuation (4 of
C 23	41	2.9	107475	45	AC017106	AC017106 Homo sapi
24	40.8	2.8	2003	5	AR022412	AR022412 Sequence
C 25	40.8	2.8	2003	5	AR068890	AR068890 Sequence
C 26	40.8	2.8	5123	16	HEPVIE	X15120 Pseudorabie
27	40.8	2.8	8438	16	SHILT	M57505 Pseudorabie
C 28	40.8	2.8	38258	10	AC000402	AC000402 Genomic S
29	40.8	2.8	44078	11	AC002102	AC002102 Homo sapi
C 30	40.6	2.8	81463	45	AC015882	AC015882 Homo sapi
31	40.6	2.8	110000	45	AC017014_1	Continuation (2 of
C 32	40.6	2.8	174707	45	AC017082	AC017082 Homo sapi
C 33	40.6	2.8	199284	33	AC006483	AC006483 Homo sapi
C 34	40.4	2.8	937	40	HOELK1	AF080615 Homo sapi
C 35	40.4	2.8	1648	9	HSELK1	Y11432 H.sapiens e
C 36	40.4	2.8	23008	9	AB016194	AB016194 Homo sapi
C 37	40.4	2.8	42655	1	SC7H2	AL109732 Streptomy
38	40.4	2.8	156601	10	HS212G6	AL009172 Homo sapi
C 39	40.4	2.8	190000	33	AC004479	AC004479 Homo sapi
C 40	40.2	2.8	3289	9	HSMFH1	Y08223 H.sapiens M
41	40.2	2.8	31636	1	SC7C7	AL031031 Streptomy
42	40.2	2.8	214258	42	AC009108	AC009108 Homo sapi
43	40	2.8	3397	9	HSDNATEF1	X84839 H.sapiens T
44	40	2.8	4443	10	HUMTEF1	M63896 Homo sapien
C 45	39.8	2.8	42210	1	SC1C2	AL031124 Streptomy

ALIGNMENTS

RESULT 1
LOCUS AF062534 2340 bp mRNA PRI
DEFINITION Homo sapiens genethonin 1 mRNA, complete cds.
ACCESSION AF062534
VERSION AF062534.1 GI:3851521
KEYWORDS

24-NOV-1998

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	495.8	43.9	771	4	US-08-972-008-3	Sequence 3, Appli
2	495.8	43.9	2525	4	US-08-972-008-1	Sequence 1, Appli
3	337.4	29.9	1753	7	5225348-2	Patent No. 5225348
4	315.2	27.8	2128	3	US-08-371-377-16	Sequence 16, Appli
5	257.4	22.8	4695	7	5225348-3	Patent No. 5225348
6	153.4	13.6	3073	1	US-07-688-352C-31	Sequence 31, Appli
7	153.4	13.6	3073	1	US-08-474-379C-31	Sequence 31, Appli
8	153.4	13.6	3073	6	PCT-US91-02714-30	Sequence 30, Appli
9	124.2	11.0	3461	4	US-08-389-504B-1	Sequence 1, Appli
10	124.2	11.0	3461	5	US-08-466-047B-1	Sequence 2, Appli
11	109.4	9.7	2095	1	US-08-329-681A-2	Sequence 14, Appli
12	39.4	3.5	7218	1	US-08-232-453-14	Sequence 3, Appli
13	38.8	3.4	2134	3	US-08-483-151-3	Sequence 34, Appli
14	38.8	3.4	2134	6	PCT-US96-06427-3	Sequence 1, Appli
15	37	3.3	453	4	US-08-387-942C-34	Sequence 12, Appli
16	37	3.3	12588	4	US-08-387-942C-1	Patent No. 5340934
17	36.4	3.2	477	7	US-08-665-647-12	Sequence 20, Appli
18	36.2	3.2	933	7	5340534-12	Sequence 21, Appli
19	35.2	3.1	246240	3	US-08-724-394A-20	Sequence 22, Appli
20	35.2	3.1	246240	3	US-08-724-394A-22	Sequence 1, Appli
21	35.2	3.1	49377	2	US-08-764-233A-1	Sequence 1, Appli
22	34.8	3.1	49377	2	US-08-403-852D-1	Sequence 1, Appli
23	34.6	3.1	5392	3	US-08-146-930-1	Sequence 1, Appli
24	34.2	3.0	6530	4	PCT-US93-03993-1	Sequence 1, Appli
25	34.2	3.0	6530	6	US-08-722-001-29	Sequence 29, Appli
26	34	3.0	1776	2	US-08-334-698-1	Sequence 1, Appli
27	34	3.0	2140	1	US-08-334-698-1	Sequence 1, Appli

	Query Match	43.9%	Score 495.8	DB 4	Length 771
	Best Local Similarity	81.1%	Pred No. 6.3e-126		
	Matches 614	Conservative	Mismatches 0	Indels 137	Gaps 6
OY	20	ctggggacactgtggccgctgtttggggagccctgtctggacagtggatccgtgggag	79		
Db	17	CAGGGCCACTGTGGCCCTGTGCCCTGGGGGGCCCTGCTGTGGGCCGTGGGCTTCCTGAGCT	76		
OY	80	ccgtgatgggtcccgaggattctgtgcccgttggtgtgtctggctccagcagccagcag	139		

Matches 133; Conservative 25; Mismatches 16; Indels 3; Gaps 3;

Db 1 MAPGVARGPTWRLRGLGAAALLLIPVAAQPPGACSONTKTCEELKNVSLWC 60
 QY 1 MAPA-NGLTLPWHMLLG-AVLLLL-SGASQAPPRVGCSEYTNRSCEELRNVSCLWC 57
 Db 61 NTKACLDYPTVSLPASCKLSSARWGVNFEALITMSVVGTLGIAICCC 120
 QY 58 NENKACMDYPTVSLPASCKLSSARWGVNFEALITMSVVGTLGIAICCC 117
 Db 121 CRKRSKPRSEKAMREERIRDEERAEKTHDEIRKYGKFKENPYARF 177
 QY 118 CRKRSKPRSEKAMREERIRDEERAEKTHDEIRKYGKFKENPYARF 174

RESULT 2
 ID A33_HUMAN STANDARD; PRT; 319 AA.
 AC Q99795;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CELL SURFACE A33 ANTIGEN PRECURSOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-COLON CARCINOMA;
 RX MEDLINE; 97165045.
 RA HEATH J.K., WHITE S.J., JOHNSTONE C.N., CATIMEL B., SIMPSON R.J.,
 RA MORITZ R.L., TU G.-F., JI H., WHITEHEAD R.H., GROENEN L.C.,
 RA SCOTT A.M., RITTER G., COHEN L., WELT S., OLD L.J., NICE E.C.,
 RA BURGESS A.W.;
 RT "The human A33 antigen is a transmembrane glycoprotein and a novel
 member of the immunoglobulin superfamily."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:469-474(1997).
 RN [2]
 RP POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE; 97396159.
 RA RITTER G., COHEN L.S., NICE E.C., CATIMEL B., BURGESS A.W.,
 RA MORITZ R.L., JI H., HEATH J.K., WHITE S.J., WELT S., OLD L.J.,
 RA SIMPSON R.J.;
 RT "Characterization of posttranslational modifications of human A33
 antigen, a novel palmitoylated surface glycoprotein of human
 gastrointestinal epithelium."
 RL Biochem. Biophys. Res. Commun. 236:682-686(1997).
 CC -1- FUNCTION: MAY PLAY A ROLE IN CELL-CELL RECOGNITION AND SIGNALING.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN NORMAL GASTROINTESTINAL
 EPITHELIUM AND IN 95% OF COLON CANCERS.
 CC -1- PTM: N-GLYCOSYLATED, CONTAINS APPROXIMATELY 8 K OF N-LINKED
 CARBOHYDRATE.
 CC -1- PTM: PALMITOYLATED.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 ONE C2-LIKE AND ONE V-LIKE DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

EMBL; U97925; AAC50957.1;
 HSP; P06907; INEU.
 MIN; 602171;
 PFAM; PF00047; Ig; 2.
 DR Immunoglobulin domain; Lipoprotein; Palmitate; Glycoprotein;
 KW Transmembrane; Signal; Antigen.
 FT SIGNAL 1 21
 CHAIN 22 319 CELL SURFACE A33 ANTIGEN.
 ~DOMAIN 22 235 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 236 256 POTENTIAL.
 FT DOMAIN 257 319 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 22 138 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 145 229 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 258 261 POLY-CYS.
 FT CARBOHYD 112 112
 FT CARBOHYD 200 200 POTENTIAL.
 FT CARBOHYD 223 223 POTENTIAL.
 SQ SEQUENCE 319 AA; 35632 MW; C4F90FC7 CRC32;

Query Match 10.6%; Score 143; DB 1; Length 319;
 Best Local Similarity 36.8%; Pred. No. 8.83e-08;
 Matches 25; Conservative 16; Mismatches 21; Indels 6; Gaps 5;
 Db 241 IAVGVVAALIIIGIIIVCC-CCKKDDNTEDKEDARENREAYEPPQLRELSREED 299
 QY 97 ITMSVLGSGVLLGITVCCYCCRRKSKRKPDKSDERAMREQ-EER--RVQREER-RAEMK 152
 Db 300 DYQEEOR 307
 QY 153 S-RHDEIR 159

RESULT 3
 ID TRHY_HUMAN STANDARD; PRT; 1898 AA.
 AC Q07283;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE TRICHOHYALIN.
 GN THH OR TRHY OR THL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93280194.
 RA LEE S.-C., KIM I.-G., MAREKOV L.N., O'KEEFE E.J., PARRY D.A.D.,
 RA STEINERT P.M.;
 RT "The structure of human trichohyalin. Potential multiple roles as a
 functional EF-hand-like calcium-binding protein, a cornified cell
 envelope precursor, and an intermediate filament-associated (cross-
 linking) protein."
 RL J. Biol. Chem. 268:12164-12176(1993).
 RN [2]
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE; 93315897.
 RA O'KEEFE E.J., HAMILTON E.H., LEE S.-C., STEINERT P.M.;
 RT "Trichohyalin: a structural protein of hair, tongue, nail, and
 epidermis."
 RL J. Invest. Dermatol. 101:65S-71S(1993).
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 DIFFERENTIATION.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 THE EPIDERMIS.
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 11:07:02 ; Search time 194.26 Seconds
(without alignments)
695.807 Million cell updates/sec

Title: US-09-276-268-10
Perfect score: 1129
Sequence: 1 ccaacatgcgcgttc.....agaatgttgaccatttt 1129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/lna/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/lna/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/lna/6_COMB.seq.*
6: /cgn2_6/ptodata/1/lna/PCUS9_COMB.seq.*
7: /cgn2_6/ptodata/1/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	495.8	43.9	771	4	US-08-972-008-3
2	495.8	43.9	2525	4	US-08-972-008-1
3	317.2	29.9	1753	7	5225348-2
4	315.2	27.8	2128	3	US-08-371-377-16
5	257.4	22.8	4695	7	5225348-3
6	153.4	13.6	3073	1	US-07-688-352C-31
7	153.4	13.6	3073	4	US-08-474-379C-31
8	153.4	13.6	3073	6	PCT-US91-02714-30
9	124.2	11.0	3461	4	US-08-389-564B-1
10	124.2	11.0	3461	5	US-08-466-047B-1
11	109.4	9.7	2095	1	US-08-329-681A-2
12	39.4	3.5	7218	1	US-08-232-463-14
13	38.8	3.4	2134	3	US-08-483-151-3
14	38.8	3.4	2134	6	PCT-US96-06427-3
15	37	3.3	453	4	US-08-387-942C-34
16	37	3.3	12588	4	US-08-387-942C-1
17	36.4	3.2	477	4	US-08-665-647-12
18	36.2	3.2	933	7	5340934-12
19	35.2	3.1	246240	3	US-08-724-394A-20
20	35.2	3.1	246240	3	US-08-724-394A-21
21	35.2	3.1	246240	3	US-08-724-394A-22
22	34.8	3.1	49377	2	US-08-764-233A-1
23	34.6	3.1	5392	3	US-08-403-852D-1
24	34.2	3.0	6530	4	US-08-146-930-1
25	34.2	3.0	6530	6	PCT-US93-03993-1
26	34	3.0	1776	2	US-08-722-001-29
27	34	3.0	2140	1	US-08-334-698-1

28	34	3.0	2140	1	US-08-228-932-1	Sequence 1, Appli
29	34	3.0	2140	2	US-08-488-939-1	Sequence 1, Appli
30	34	3.0	2140	3	US-08-406-855A-1	Sequence 1, Appli
31	34	3.0	2140	4	US-08-722-190-1	Sequence 1, Appli
32	34	3.0	2140	6	PCT-US95-04203-1	Sequence 1, Appli
33	34	3.0	2363	2	US-07-923-724-7	Sequence 7, Appli
34	34	3.0	2363	3	US-08-609-426A-7	Sequence 7, Appli
35	34	3.0	2379	3	US-08-374-652C-1	Sequence 1, Appli
36	34	3.0	4897	7	5196516-7	Patent No. 5196516
37	34	3.0	6049	6	PCT-US95-11684-3	Sequence 3, Appli
38	33.8	3.0	4257	3	US-08-690-473-1	Sequence 1, Appli
39	33.8	3.0	12001	2	US-08-458-568A-11	Sequence 11, Appli
40	33.6	3.0	44377	3	US-08-387-942C-30	Sequence 30, Appli
41	33.6	3.0	44377	3	US-08-804-227C-7	Sequence 7, Appli
42	33.6	3.0	44377	4	US-08-804-198-1	Sequence 1, Appli
43	33.4	3.0	450	4	US-08-387-942C-32	Sequence 32, Appli
44	33.4	3.0	1212	7	5212296-17	Patent No. 5212296
45	33.4	3.0	1497	3	US-08-820-170A-41	Sequence 41, Appli

ALIGNMENTS

RESULT 1
US-08-972-008-3
; Sequence 3, Application US/08972008
; Patent No. 5942420
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: No. 5942420el Molecules of the Follistatin-Related
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,008
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 771 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-972-008-3

Query Match 43.9%; Score 495.8; DB 4; Length 771;
Best Local Similarity 81.1%; Pred. No. 6.3e-126;
Matches 614; Conservative 0; Mismatches 137; Indels 6; Gaps 3;
QY 20 ctggggcactgtggcgcctgtttggggagccctgtgtgacagtggatccgtggcg 79
Db 17 CAGGGCCACTCTGCCCTCTGCCCTGGCGGCGCTTGGCGGCTTGGCTTCTGAGCT 76
QY 80 ccgtgatgggctccagagattctgtgcccgtggtgctgctccagcagggcagag 139

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 11:00:19 ; Search time 3493.2 Seconds
(without alignments)
880.902 Million cell updates/sec

Title: US-09-276-268-9
Perfect score: 815
Sequence: 1 aggtgacactagtggatcc.....gtagtataaacaagtcacgc 815

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

77 36
4.2

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- EST:*
- 1: em_est1:*
 - 2: em_est2:*
 - 3: em_est3:*
 - 4: em_est4:*
 - 5: em_est5:*
 - 6: em_est6:*
 - 7: em_est7:*
 - 8: em_est8:*
 - 9: em_est9:*
 - 10: em_est10:*
 - 11: em_est11:*
 - 12: em_est12:*
 - 13: em_est13:*
 - 14: em_est14:*
 - 15: em_est15:*
 - 16: em_est16:*
 - 17: em_est17:*
 - 18: em_est18:*
 - 19: em_est19:*
 - 20: gb_est1:*
 - 21: gb_est2:*
 - 22: gb_est3:*
 - 23: gb_est4:*
 - 24: gb_est5:*
 - 25: gb_est6:*
 - 26: gb_est7:*
 - 27: gb_est8:*
 - 28: gb_est9:*
 - 29: gb_est10:*
 - 30: gb_est11:*
 - 31: gb_est12:*
 - 32: gb_est13:*
 - 33: gb_est14:*
 - 34: gb_est15:*
 - 35: gb_est16:*
 - 36: gb_est17:*
 - 37: gb_est18:*
 - 38: gb_est19:*
 - 39: gb_est20:*
 - 40: gb_est21:*
 - 41: gb_est22:*
 - 42: gb_est23:*
 - 43: gb_est24:*
 - 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss12:*
- 103: gb_gss13:*
- 104: gb_gss14:*
- 105: gb_gss15:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	550.2	67.5	591	46	AI466595 mg73d12.y
2	466.8	57.3	470	44	AI287088 ui79e01.y

517 TATTGACAACTGTACGGTTTCAGCAACTACTCATCCACCAAGGCAAGGATCCAAAGT 576

RESULT	2	EST	24-NOV-1998
AI287088	470 bp	mus	
LOCUS	AI287088	mouse	
DEFINITION	u175e01.v1	liver	
IMAGE:188248.5'	AI287088	musculus	
ACCESSION	AI287088	sequence.	
VERSION	AI287088.1		
KEYWORDS	GI:3926841		
EST.			

SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 470) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lucy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennnon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 19, 1998 this sequence version replaced g1:2287581.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:972572

```

seq primer: custom primer used
High quality sequence stop: 453.
Location/Qualifiers
1..470
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1888248"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pWE18S-FL3; Site.1: DraIII
(CACTGTGTG); Site.2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGCCCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGTGGCTACTGG], digested
and cloned into distinct DraIII sites of the pWE18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGCG and 3' end
primer CGACCTCTGACCTCGGACGAC."
117 a 145 c 120 g 88 t

```

```

Query Match          57.3%; Score 466.8; DB 44; Length 470;
Best Local Similarity 99.6%; Pred. NO. 2e-105;
Matches 468; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 56 gtgcgcggagagcagctctgtggcgccggcggtgtgtcggggcgctcagggcgccctatqqa 115

```

1	GTGGCGCGAGAGACAGCTCTCTGGGCGCGCGGCGGTGCTGGGCGCGCTCAGGGGCGCCCTGGGA	60
116	acaatggcgctgtgcgcggcgcgctgtcgtcgtcggcgctgtcgcagggtgcgtggcgctg	175
61	ACAATGGCGCTGTGCGCGCGCGCGGCTGTCTGCTGGCGGTGCTGTGCAGGTGCTGTGGCGCTG	120
176	ctaggggcgcgcgagacccgacgcgtcagggtctcgaagtgcgaatggaaaccactcagt	235
121	CTAGGGCGCGGAGGACCCGCCACGCTCAGGGCTCTGCAAGTGTGAAACCACTCAGTG	180
236	ctgacctccaattataacataacagagaataccaaaccagacacatgagtgtgtttccaac	295
181	CTGACCTTCCAATATTACATATACAGAGAAATACCAACACGACCATGAGTGTGTTCACAC	240
296	cagaccagtgaatgcagagcaccgcgaagcctccgtactgcgaataaacacacacatt	355
241	CAGACCAAGTGAATATGAGAGCACCGGGAAGCCTTCGGTACTGCCAAAACATCCACACATT	300
356	atcactgtaaacctgcaactatgttataaatatacaacccaggagcttaccacatgtg	415
301	ATCACTGTGAACCTGCACTATTGTAAATAATCAACCCGAGAGTCTTACTACATGTG	360
416	acgcctactgcctcaagtctcacaccaatgcgaagtgtcttcacaaactaccaccaag	475
361	ACGCCTACTGCCTAAGTCTACACCAATGCAGTGTCTTCTCAAACTCTAACCACAG	420
476	tcagcatcccatgacaacccccacccacagtagttattgacaactgtaac	525
421	TCAGCATCCATGACAACCCCGACACAGTAGTATTATTGACAACGTGAAC	470

RESULT 3

AA998505	AA998505	473 bp	mrna	EST	04-JUL-1999
LOCUS	UI-R-C0-1e-g-11-0-UI.s1	UI-R-C0	Rattus norvegicus	cdna clone	
DEFINITION	UI-R-C0-1e-g-11-0-UI 3',	mrna sequence.			
ACCESSION	AA998505				
VERSION	AA998505.1	GI:4290322			
KEYWORDS	EST.				
ORGANISM	Norway rat.				
	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 473)				
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res. 6 (9), 791-806 (1996)				
MEDLINE	97044477				
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189156.				

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9363
Email: msoares@blue.weeg.uiowa.edu

Oligo-dT track not found, Not I si

is likely internal to the message.

FATIMA BONAIDO, PH.D. Clone distri-
through Research Genetics. This clo-

I.M.A.G.E. Consortium at LLNL (inf

ID=1782202

Seq primer: M13 Forward
DQ1YA=No

FEATURES	Location/Qualifiers
FOUR-NO.	

source 1. .473

```
/organism="Rattus norvegicus"
/attr={n="Species-Name"}
```

```
/scran= sprague-dawley-  
/db xref="taxon:10116"
```

```

/clone="UI-R-C0-1e-g-11-0-UI"

```

```
./clone_lib="UI-R-C0"
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 11:00:01 ; Search time 3493.2 Seconds
(without alignments)
1222.454 Million cell updates/sec

Title: US-09-276-268-7
Perfect score: 1131
Sequence: 1 ggcaccagccgcgtcttgt.....aaagaccactggaagcgaa 1131

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	553.8	49.0	591	41	AI050489	AI050489 uc86g10.y
2	537.8	47.6	570	28	AA116725	AA116725 mq24g04.f

3 531.4 47.0 578 40 C88549 Mous
4 503.4 44.5 521 47 A1526757 u142e07.Y
5 487 43.1 527 28 A116721 mq24e04.r
6 486 43.0 508 27 AA050029 mj39d12.r
7 469.8 41.5 525 41 A1043112 u1043112
8 462.8 40.9 508 34 AA475668 v122f12.r
9 447.4 39.6 530 33 AA388260 vb59f08.r
10 440.4 38.9 453 40 AA967207 ua50g04.r
11 436.8 38.3 450 41 A1049344 u1049344
12 433 38.6 504 31 AA285576 vb83c12.Y
13 425.4 37.6 483 27 W81980 me96c02.r1
14 421 37.2 443 39 AA880246 v199g07.r
15 414 36.6 436 26 W98064 me73c04.r1
16 411 36.3 425 33 AW012675 u184b10.Y
17 409.4 36.2 433 36 AA647122 ve44b09.r
18 399.8 35.3 517 47 A1528172 u194c07.Y
19 399.6 35.3 450 34 AA500267 v197e03.r
20 396.2 35.0 494 38 AA800515 EST190012
21 396.2 35.0 503 43 A1101122 EST210411
22 388.8 34.4 499 43 A1237311 EST233873
23 388 34.3 497 40 AA944801 EST200300
24 386 34.1 508 45 AA818341 UI-R-A0-a
25 382 33.8 393 35 AA546047 vk6ic12.r
26 374.2 33.1 530 69 AW140424 EST290400
27 373.2 33.0 410 42 A1153013 ue60b11.r
28 371.2 32.8 492 23 H34931 EST106045 R
29 366.4 32.4 475 69 AW140753 EST290748
30 366.2 32.4 502 69 AW140420 EST290396
31 357.8 31.6 475 45 AA859375 UI-R-E0-b
32 355.8 31.5 452 49 AV008016 AV008016
33 355.2 31.4 395 43 A1180883 ub86f05.r
34 353 31.2 374 26 W98121 mg25g04.r1
35 345.4 30.5 358 26 W43944 mc72b08.r1
36 344.4 30.5 378 26 W47820 mc90g02.r1
37 342 30.2 342 25 W10617 ma39d12.r1
38 341 29.8 394 26 W62867 md85g08.r1
39 337.4 29.8 350 26 W57201 md59a11.r1
40 337.4 29.8 439 38 AA800516 EST190013
41 336 29.7 380 33 AA413130 ve97f01.r
42 333.8 29.5 339 39 AA840217 ud01n06.r
43 332.2 29.4 430 39 AA851829 EST194597
44 329 29.1 388 33 AA413131 ve97f02.r
45 328.8 29.1 332 28 AA073175 mm94b11.r

ALIGNMENTS

RESULT 1
LOCUS A1050489 591 bp mRNA EST 09-JUL-1998
DEFINITION uc86g10.y1 Sugano mouse kidney mkiia Mus musculus cDNA clone
IMAGE:1432578 5' similar to WP:F49C12.12 CE03372 ; mRNA sequence.
ACCESSION A1050489
VERSION A1050489.1 GI:3299606
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 591)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1998)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043618.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.vustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:916646
Seq primer: custom primer used
High quality sequence stop: 510.
FEATURES
source
1..591
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1432578"
/clone_lib="Sugano mouse kidney mkiia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pME18S-FL3; Site1: DraIII
(CACTGTGNG); Site2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ARGTGGCCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGNG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAGCTGG and 3' end
primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 125 a 190 c 133 g 143 t
ORIGIN
Query Match 49.0%; Score 553.8; DB 41; Length 591;
Best Local Similarity 99.6%; Pred. No. 9.8e-118;
Matches 555; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 9 ccgggtcttgctgcgtcagtcctcagcgagtcctcctcctccctccatgcgc 68
Db 18 CCCGGCTTCGTGCTCGGCTCAGTCCTCCAGCGATTCTCCTACCTCCGCCCTCCATGCG 77
QY 69 gtgcctctgtgctggtggcctaagctggcgccctgtggcctcctcagcgctgggg 128
Db 78 GTGGCTCTGTGCTGTGGGCTAAGCTGGCGGCTGTGGCATGCTCCTCAGCGCTGGGG 137
QY 129 agtgatcatgtgataatgctcgggataattttcaatgtccattctgtgttaattga 188
Db 138 AGTGATCATGTTGATGATGCTCGGATATTTTCAATGTCCTGCTGTGTTAATGA 197
QY 189 ggagcttccctccacagagaagattttgagacggtcctcagaaacatatacaacctga 248
Db 198 GGAGGTTCCTTCCACAGAGAAAGATTTTGAGACGCTCCTCAGAACATATACACCTGTA 257
QY 249 cgagcaagtgatgataactgttttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 308
Db 258 CGAGCAAGTCAGCTACAACTGTTTCATCGCGGGGCTCTACCTCCTCCTCGGAGGCTT 317
QY 309 ctctctctgccaagtctcgtctcaacaagcgaggaatacatgtgctagagcgcggt 368
Db 318 CTCCTTCTGCCAAGTTCGTCTCAACAAGCGCAAGGAATACATGGTGGCTAGAGCGCGT 377
QY 369 ccgctctcctcccccagccccctctctatttaagacctccgagacctccctccacac 428
Db 378 CGCCCTCTCCCTCCCGAGGCCCTCTCTATTAAAGATCCGAGACTCCCTCCACATC 437
QY 429 atctggtcctcttgggactgtgacctagcgagacgtcatccctggcctggcaaac 488
Db 438 ATCTGGCGTCTTTGGGACTTGTGACCTTAGCGAGAGCTATCCTCGGCGCTGCAAACT 497
QY 489 ggcggcagctctggaggagacgggggacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 548

Db	79	GTGCGTCTGTGCTGTGGCGCTAAGCTGCGCGCGCTGTGG-ATCGTCCCTCAGCGCGCTGGGG	137
Oy	129	agtgatcatgttgataatgctcgggatatttttcaatgtccattctgctggttaattga	188
Db	138	AGTGATCATGTTGATTAATGCTTCGGGATATTTTTCATGTGCCATTCTGCTGTGTTAAATTGA	197
Oy	189	ggacgttcccttcacagagaaagattttgagaacggtctctcagaaacatatcaaacctgta	248
Db	198	GGACGTTCCCTTCACAGAGAAAGATTTTGAGAACGGTCTCTCAGAACATATACACCTGTGA	257
Oy	249	cgagcaagtgcagctcaaacgttttcatctgcgcgcgggacctctacctctctcgagagctt	308
Db	258	CGAGCAAGTCAGTCAACTGTTTTCATCGCGCGGGCCTCTACCTCTCTCGGAGGCTT	317
Oy	309	ctccttctgccaagtctgctctcaacaaacgcgaaggaatacatggtgcgtctagagcgcggt	368
Db	318	CTCCTTCTGCCAAGTTCGTCCTCAACAACGGCAGAGATATCATGTGCGCTAGAGCCCGGT	377
Oy	369	cgcctctccctccccagccccctctctattttaaaactccgcagactcgtcccaactc	428
Db	378	CGCCCTCTCCCTCCCGACGCCCTTCTCTATTAAAGACTCCGCAGACTCCGTCCTCCACCT	437
Oy	429	atctggcgctcttctggagcttgtaacctagcagagagctcatccctggccctgcaaaact	488
Db	438	ANTCTGGGTCCTTTGGGACTTGTGACCCTAGCAGAGCTCATCCCTGGCCCTGCAAAACT	497
Oy	489	gccccagcctctggagagaccagggtgaccgcgcgccttctgaactacaataaaaa	548
Db	498	GGCCCAAGCTCTGGAGGAGACCGAGGGTGACCGCGCCCGCTTCTGAATCAATAAAAA	557
Oy	549	gaagcgggtcccc	561
Db	558	GAAGCGGTTCGCC	570
RESULT	3		
C88549			
LOCUS		578 bp	EST
DEFINITION	C88549	Mouse early blastocyst cdNA Mus musculus	28-MAY-1998
ACCESSION		08007610X00, mRNA sequence.	
VERSION	C88549		
KEYWORDS			
SOURCE	C88549.1	GI:3164311	
ORGANISM		house mouse.	
REFERENCE		Mus musculus	
AUTHORS		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE		1 (bases 1 to 578)	
JOURNAL		Sasaki, N., Nagaoka, S., Itoh, M., Izawa, M., Konno, H., Carninci, P., Yoshiki, A., Kusakabe, M., Moriuchi, T., Muramatsu, M., Okazaki, Y., Hayashizaki, Y.	
MEDLINE		Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones	
COMMENT		Genomics 49 (2), 167-179 (1998)	
		98260665	
		On Nov 29, 1993 this sequence version replaced gi:637711.	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 24, 2000, 10:42:11 ; Search time 5166.57 Seconds
(without alignments)
-843.347 Million cell updates/sec

Title: US-09-276-268-6
Perfect score: 1435
Sequence: 1 catggcgccgtctgtgctag.....acacagctaaagcaacactg 1435

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: gb_ba1.*
37: gb_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	582.6	40.6	2340	40	AF062534 Homo sapi
2	429.8	30.0	113457	11	AC003686 Homo sapi
3	142	9.9	339	13	G37764 SHGC-63501
C 4	57	4.0	7218	5	I66494 Sequence 14
5	46	3.2	159515	12	AC002327 Mus muscu
C 6	45.8	3.2	3281	34	LMA243459 Leishmani
7	45	3.1	96106	45	AC017015 Homo sapi
C 8	43.6	3.0	150079	41	AC009447 Homo sapi
9	42.4	3.0	2685	10	HUMBA3
10	42.4	3.0	43058	9	HSGG1
C 11	42.2	2.9	742	40	AF000672 Homo sapi
C 12	42.2	2.9	2266	9	HUMELK1A
C 13	42.2	2.9	29358	34	DMBH6115
14	42.2	2.9	77656	45	AC017522 Drosophil
C 15	41.6	2.9	25306	2	AF058302 Drosophil
C 16	41.6	2.9	56414	1	MTV002 Mycobacte
17	41.6	2.9	149977	10	CNS01DSV
18	41.6	2.9	174707	45	AC017082 Homo sapi
C 19	41.6	2.9	232426	33	HS1086L22
20	41.4	2.9	202776	41	AC007871 Homo sapi
C 21	41	2.9	42833	32	HS367G8
22	41	2.9	100516	32	HSS171M_3
C 23	41	2.9	107475	45	AC017106 Continuation (4 of
24	40.8	2.8	2003	5	AR022412 Sequence
25	40.8	2.8	2003	5	AR068890 Sequence
C 26	40.8	2.8	5123	16	HEPVIE
27	40.8	2.8	8438	16	SHILLT
C 28	40.8	2.8	38258	10	AC000402 Pseudorabie
29	40.6	2.8	44078	11	AC002102 Genomic s
30	40.6	2.8	81463	45	AC016882 Homo sapi
C 31	40.6	2.8	110000	45	AC016882 Homo sapi
C 32	40.6	2.8	174707	45	AC017082 Homo sapi
33	40.6	2.8	199284	33	AC006483 Homo sapi
C 34	40.4	2.8	937	40	HOELK1
C 35	40.4	2.8	1648	9	HSELK1
C 36	40.4	2.8	23008	9	AB016194 Homo sapi
C 37	40.4	2.8	42655	1	SC7H2 Streptomy
38	40.4	2.8	156601	10	HS21266 Homo sapi
C 39	40.4	2.8	190000	33	AC004479 Homo sapi
C 40	40.2	2.8	3289	9	HSMFHI
41	40.2	2.8	31636	1	SC7C7 Streptomy
42	40.2	2.8	214258	42	AC009108 Homo sapi
43	40	2.8	3397	9	HSDNATEF1
44	40	2.8	4443	10	HUNTEF1
C 45	39.8	2.8	42210	1	SC1C2 Streptomy

ALIGNMENTS

RESULT 1
AF062534
LOCUS Homo sapiens genethonin 1 mRNA, complete cds.
DEFINITION AF062534
ACCESSION AF062534.1 GI:3851521
VERSION
KEYWORDS

24-NOV-1998

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2340)
 AUTHORS Pietu,G., Allibert,O., Guichard,V., Lamy,B., Bois,F., Leroy,E.,
 Mariage-Sampson,R., Houllgate,R., Soularue,P. and Auffray,C.
 TITLE Novel gene transcripts preferentially expressed in human muscles
 revealed by quantitative hybridization of a high density cDNA array
 JOURNAL Genome Res. 6 (6), 492-503 (1996)
 MEDLINE 96425696
 REFERENCE 2 (bases 1 to 2340)
 AUTHORS Bouju,S., Lignon,M.F., Pietu,G., Le Cunff,M., Leger,J.J.,
 Auffray,C. and Dechesne,C.A.
 TITLE Molecular cloning and functional expression of a novel human gene
 encoding two 41-43 kDa skeletal muscle internal membrane proteins
 JOURNAL Biochem. J. 335 (Pt 3), 549-556 (1998)
 MEDLINE 99013693
 REFERENCE 3 (bases 1 to 2340)
 AUTHORS Bouju,S., Lignon,M.F., Pietu,G., Le Cunff,M., Leger,J.J.,
 Auffray,C. and Dechesne,C.A.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAY-1998) Unit 300, INSERM, 15 Avenue Charles
 Flahault, Montpellier 34060, France
 FEATURES
 source
 1..2340
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4q21-q25"
 /clone="GENX-3414"
 /tissue_type="muscle"
 128..1204
 /note="contains leucine zipper motif"
 /codon_start=1
 /product="genethonin 1"
 /protein_id="AAC78827.1"
 /db_xref="GI:3851522"
 /translation="MGAVNSALLVGGGLAGALFVLLRGGPDGDKGDAEKEKDALP
 GGAIPGGHQSGSLSPGSGOELVTPKPEHLSNGLHSKIDKLGKLAASWRLQ
 PSREKATDSHPVQPDTEPATSETSNRSYSSEVRSNLESLESPGWEVFOKGOE
 ISAKATCFAPKLPSSNLLKNFAKEMSLSDLSNODRDVHEWVPHBSWGDVGVG
 GSKLPVNLNQGNDNGRSTLVARGQQVHGKRVAVNPAGSQVSVRFVHYTST
 DVQFLAVTGHEDHGRWNTYIPLHYNKDGFNSHFLPADTVVWENKFLVLENGVTRW
 EECNSRFLETGHEDKRVHAWMGIH"
 BASE COUNT 642 a 487 c 619 g 592 t
 ORIGIN

QY 301 ggc-----gctgcagaatgttgagcagcagctgggtcctaatgccagagagattgttctcgt 354
 Db 418 ATGGAGATTGCAGATCCCTTCCAGGGAAGCTGTGACAAATTCAGAGAACATGTTCTTC 477
 QY 355 tgggaagattccagacaca-----caactccagggccga 387
 Db 478 TGGACAGTTTCCAGACACACAGAGAGTCCAGCTACCTCTGTAGACAGTAACCTCTAGGATTA 537
 QY 388 cctcgaagcggcgaagaatacaagcccgagatctcatggaggagagatggagactcccaaa 447
 Db 538 CTCTGAAGTTTCAAGAAATGAAGCCTTGAATCTCCTATGGAGAAATGGGATTCACAAA 597
 QY 448 aggaacaagaaacagctgtcaasgtagctggcagtggtgcgcgcaagctggcctccagcag 507
 Db 598 AGGACAAGAGATATCTGCTTAAGCAGCTACATGTTTGGCAGAGAAATGGCTTCTAGCNA 557
 QY 508 cctgcttggcagagagctaaagcag-----taagtccaggaaca 546
 Db 658 CTGCTCAAGAACAGAGAGCTAAAGAGAAATGAGCCTCTCTGATTTGAACAGTCAGGACCG 717
 QY 547 ggcagccacagagagctgggaagtggtcttaggcaactcatcttggggagtggtgttt 606
 Db 718 GGTGACACAGAGAGTGGGAAATGCTTGGCTTAGGCACTCATCTTGGGGGATGTTGGTGT 777
 QY 607 ggtggcagcttggcgttcttagtttaagtctaaatcagagaatggacagacacaaa 666
 Db 778 GGTGGCAGCTTTAAGGCTCCAGTGTTAAACCTAAACAGGGAATGGCAATGGGAAGAG 837
 QY 667 cagcttgggaggaagaggtgggaagtatagatgggaagtggca-----713
 Db 838 CACTTGGTGGAGAGAGAGGTGAGCAAGTGCATGGGAAATGGAAAGGTAGCAGTGAT 897
 QY 713 -tcttgaaacctcaacaggtcagcatccagttccaggtccagctacacacacacacacoga 771
 Db 898 GCCTGCAGGGCTCTCAGCAAGATTAGTGTCCAGGTCCAGTCCATTTGTTCACAAAGCAGTGA 957
 QY 772 tgtcagctcattcagtagtactggagacacatgagaccttgggagatggaacacacacat 831
 Db 958 TGTGCAATTCATTGCAAGTAACTGGAGACCATGAGTGTCTTGGAGATGGAACACTTACAT 1017
 QY 832 cccactccactactgcaagagcggctcgtggtctcattctgtctctcctcctcctgcagacac 891
 Db 1018 CCACATCCACTATACAAAGATGGGTCTGGCTCATTCATTCCTCTCCTGCCGAGATAC 1077
 QY 892 agtgggtggagtggaagtctggttgtagagataaaggaagtactctgttgggaagaatg 951
 Db 1078 AGTGGTGGAGTGGAAAGTTTGTGTGTGTAGAGAAATGGGGAGTTTACCCGCTGGGAAGAATG 1137
 QY 952 cagcaatagattcctcagactggccatgagataaagtggttcattcgttgggtgggagat 1011
 Db 1138 CAGCAATAGATTCTTAGAAACTGGCCATGAGGATAAAGTGGTTCACGCAATGGTGGGGAT 1197
 QY 1012 tcatcactcagtttcagagcatccaagcgtcgcagcagaaatggaagagctgaag 1071
 Db 1198 TCACTGATTCAGTTTCAAGATTAATGGAGAAAGCTGTAGAACAAATGTGAAGAGGCTGAGG 1257
 QY 1072 ctttagagcgcactgcatgacttaagagcgggtggtgattcccaattgtgacctag 1131
 Db 1258 TTGTGGAACACACTGAA-----TAAATAAAGGCAGTGTGACTCCAAATTCAGCCATCG 1312
 QY 1132 ggtctttcagatttctgtagtggtgtttgttcccaaatgttaggaagatgtatgct 1188
 Db 1313 AATTGTTTAAATTT---GCTAGTGGATTTTGTCTACTGTGCAGAAATATATATGTCT 1366

RESULT 2
 AC003686
 LOCUS Homo sapiens 113457 bp DNA PRI 09-JUN-1998
 DEFINITION Human PAC library) complete sequence.
 AC003686
 ACCESSION AC003686
 VERSION AC003686.1 GI:3201599
 KEYWORDS HTG.

17:27
 SOURCE ORGANISM
 REF:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 10:59:29 ; Search time 3493.2 Seconds
(without alignments)
569.053 Million cell updates/sec

Title: US-09-276-268-3
Perfect score: 619
Sequence: 1 ggccaccagggaagccctgcc.....tgtgcccaggcccaactta 619

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues 9077268
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss12:*
- 103: gb_gss13:*
- 104: gb_gss14:*
- 105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	420.8	68.0	588	39	C86502
C 2	190.4	30.8	576	30	AA206019
					C86502 C86502 Mous
					AA206019 zq54d06.s

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 10:59:23 ; Search time 3493.2 Seconds
(without alignments)
744.713 Million cell updates/sec

Title: US-09-276-268-2
Perfect score: 689
Sequence: 1 gtcgcctgaggtcccccgcg.....tgatccctcagcctctactg 689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss12:*
- 103: gb_gss13:*
- 104: gb_gss14:*
- 105: gb_gss15:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	575	83.5	576	42	A1119658	A1119658 uf05f04.y
2	556	80.7	560	60	A1790880	A1790880 uk29b06.y


```

Db 482 GGAATGAAGTCAAGCATCATGAATCAGAAAAAATACGGTCTGTTTAAAGAACAAAA 541
Qy 541 cccgtatgagaagtctaaagtgctggcgccacac 575
Db 542 CCGGTATGAGAAGTCTTAAGTGGCTGGCACAC 576

RESULT 2
AI790880
LOCUS
DEFINITION
AI790880 560 bp mRNA EST 02-JUL-1999
IMAGE:1970387 5' similar to SW72111_HUMAN P53801 PUTATIVE SURFACE
GLYCOPROTEIN C21ORF1 PRECURSOR ; , mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 560)
AUTHORS
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE
The WashU-NCI Mouse EST Project 1999
JOURNAL
Unpublished (1999)
COMMENT
On May 18, 1998 this sequence version replaced gi:3137684.
Other_ESTs: uk29D06.x1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:987127
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 525.
FEATURES
Location/Qualifiers
1..560
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1970387"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: Draili
(CACCATGTG); Site_2: Draili (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCGCTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a Draili adaptor [GTGTGGCTACTGG], digested
vector [5' site CACTGTGTC, 3' site CACCATGTG]. XhoI shou
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAGACTGCG and 3' end
primer CGACCTCGAGTCGAGCACA."
BASE COUNT 127 a 139 c 178 g 116 t
ORIGIN

Query Match 80.7%; Score 556; DB 60; Length 560;
Best Local Similarity 100.0%; Pred. No. 1.8e-133;
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 10:07:12 ; Search time 5166.57 Seconds
(without alignments)
-404.924 Million cell updates/sec

Title: US-09-276-268-2
Perfect score: 689
Sequence: 1 gtcgcctgaggtcccgccg.....tgatccctcagctctactg 689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pr1.*

10: gb_pr2.*

11: gb_pr3.*

12: gb_ro.*

13: gb_sts.*

14: gb_sy.*

15: gb_un.*

16: gb_v1.*

17: gb_v2.*

18: em_fun.*

19: em_hum1.*

20: em_hum2.*

21: em_in.*

22: em_or.*

23: em_ov.*

24: em_pat.*

25: em_ph.*

26: em_pl.*

27: em_ro.*

28: em_sts.*

29: em_sy.*

30: em_un.*

31: em_v1.*

32: gb_htg1.*

33: gb_htg2.*

34: gb_in1.*

35: gb_in2.*

36: em_ba1.*

37: em_ba2.*

38: em_hum3.*

39: em_hum4.*

40: gb_pr4.*

41: gb_htg3.*

42: gb_htg4.*

43: gb_htg5.*

44: gb_htg6.*

45: gb_htg7.*

46: em_htg1.*

47: em_htg2.*

48: em_htg3.*

49: em_hum5.*

50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	327	47.5	2617	9	HSSGPN15	250022 H.sapiens m
2	158	22.9	10687	43	AC015892	AC015892 Mus muscu
3	150.2	21.8	2230	10	HUMCTGREP	L48984 Homo sapien
4	118.8	17.2	17873	44	AC011782	AC011782 Homo sapi
5	118.8	17.2	33303	9	HSS171	AJ011930 Homo sapi
C 6	95.8	13.9	10687	43	AC015892	AC015892 Mus muscu
C 7	59.2	8.6	7218	5	I66434	I66494 Sequence 14
C 8	47.4	6.9	193387	11	AF064858	AF064858 Homo sapi
C 9	45.2	6.6	37655	40	AC007792	AC007792 Homo sapi
C 10	44	6.4	181098	11	AC004216	AC004216 Homo sapi
C 11	43.8	6.4	190837	11	HS9016	297353 Human DNA s
C 12	43.6	6.3	122640	41	AC009401	AC009401 Homo sapi
C 13	43.6	6.3	281632	41	AC011407	AC011407 Homo sapi
C 14	43.4	6.3	81868	41	AC009689	AC009689 Homo sapi
C 15	43.4	6.3	105607	43	AC013410	AC013410 Homo sapi
C 16	43.2	6.3	2969	12	MUSPL10	M60847 Mouse lipop
C 17	43.2	6.3	93145	43	AC012414	AC012414 Homo sapi
C 18	43.2	6.3	108064	41	AC010404	AC010404 Homo sapi
C 19	43	6.2	213581	42	AC009764	AC009764 Homo sapi
C 20	42.8	6.2	645	12	MMU239082	AJ239082 Mus muscu
C 21	42.6	6.2	160783	12	AC009287	AC009287 Mus muscu
C 22	42.4	6.2	379	13	HUMUT953	L18481 Human chrom
C 23	42.4	6.2	60332	43	AC012538	AC012538 Homo sapi
C 24	42.4	6.2	121978	43	AC016033	AC016033 Homo sapi
C 25	42.2	6.1	34612	44	AC016289	AC016289 Homo sapi
C 26	42.2	6.1	189906	33	HSJ738A13	AL109801 Homo sapi
C 27	41.8	6.1	4163	9	HSGARPGNA	Z24680 H.sapiens g
C 28	41.8	6.1	148340	42	AC011061	AC011061 Homo sapi
C 29	41.6	6.0	35000	44	AC016528	AC016528 Leishmani
C 30	41.6	6.0	168468	40	AC005510	AC005510 Homo sapi
C 31	41.6	6.0	175897	42	AC010723	AC010723 Homo sapi
C 32	41.4	6.0	317	13	G09243	G09243 human STS C
C 33	41.4	6.0	142075	40	AC005519	AC005519 Homo sapi
C 34	41.2	6.0	4238	12	RNCNG41	AJ000496 Rattus no
C 35	41.2	6.0	161987	40	AC006070	AC006070 Homo sapi
C 36	41	6.0	5166	12	AF122014	AF122014 Mus muscu
C 37	41	6.0	43900	11	AC005265	AC005265 Homo sapi
C 38	41	6.0	67841	44	AC016096	AC016096 Homo sapi
C 39	41	6.0	118195	40	AC007899	AC007899 Homo sapi
C 40	41	6.0	122383	11	HSJ944N18	AL049738 Human DNA
C 41	41	6.0	151490	45	AC011883	AC011883 Homo sapi
C 42	40.8	5.9	71342	43	AC015833	AC015833 Homo sapi
C 43	40.6	5.9	55520	11	HSJ488M4	AL078598 Human DNA
C 44	40.6	5.9	77516	10	AC002378	AC002378 Human PAC
C 45	40.6	5.9	129968	10	HS249F5	AL035691 Human DNA

ALIGNMENTS

RESULT 1
HSSGPN15
LOCUS HSSGPN15 2617 bp mRNA PRI
DEFINITION H.sapiens mRNA for surface glycoprotein.
ACCESSION Z50022
VERSION Z50022.1 GI:1107702
KEYWORDS surface glycoprotein.

30-APR-1998

SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS

Yaspo,M.L., Aaltonen,J., Horelli-Kuitunen,N., Peltonen,L. and
Lehrach,H.

TITLE

Cloning of a novel human putative type Ia integral membrane protein
mapping to 21q22.3

JOURNAL

Genomics 49 (1), 133-136 (1998)

MEDLINE

98234553

REFERENCE
AUTHORS

YASPO,M.

TITLE

Model for a transcript map of human chromosome 21: isolation of new
coding sequences from exon and cDNA libraries

JOURNAL

Unpublished

REFERENCE
AUTHORS

YASPO,M.

TITLE

Direct Submission

JOURNAL

Submitted (07-JUL-1995) Marie-Laure Yaspo, Genome Analysis,
Imperial Cancer Research, Fund, 44 Lincoln's Inn Fields, London,
WC2A 3PX, UK

FEATURES

Location/Qualifiers

1..2617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21q22.3"
/clone="HTECDNA1N15"
/dev_stage="fetus 21 weeks"
/tissue_type="thymus"
/clone_lib="directionally cloned poly(dT)-primed cDNA
library"
1..93
1..2617
94..636
/note="complete cDNA sequence for a putative plasma
integral protein containing an internalisation signal in
the cytoplasmic C-terminal domain"
/citation=[1]
/codon_start=1
/product="putative surface glycoprotein"
/protein_id="CAA90325.1"
/db_xref="GI:1107703"
/db_xref="SWISS-PROT:P53801"
/translation="MAPGVARGPTPYWRLRGLGGAALLLLIPVAAQPPPGAACSONT
NKTCBECKNVSLMCNTAKLDYPVTSLPAPSAKLSSARWGVCVNVEALIITM
SVVGTLILGIAICCCCRKRKRPSEKAWEERIRIQEERRAEKTRHDE
IRKKYGLFKEENPYAFENN"

sig_peptide
3'UTR
polyA_signal
polyA_site

BASE COUNT 625 a 672 c 673 g 647 t

ORIGIN

Query Match 47.5%; Score 327; DB 9; Length 2617;
Best Local Similarity 78.5%; Pred. No. 1.7e-76;
Matches 406; Conservative 0; Mismatches 105; Indels 6; Gaps 1;

QY 49 ccttgaggtagcgcgacactgggtgatgctctcggtgcg-----tgctgctgttgt 102
|| || || || || || || || || || || || || || || || || || || || ||
Ddb 111 CGCGGGCGGACGCCGTACTGGAGTTGGCCTCGTGCGCGCGCTGCTCTGCTGCT 170
|| || || || || || || || || || || || || || || || || || || || ||
QY 103 tcgtccgagacctccgcgcagaaacctccgagatgggttctctgagtacacaaaag 162
|| || || || || || || || || || || || || || || || || || || || ||
Ddb 171 CATCCCGTGGCGCGCGCACGAGCCTCCCGAGCTGCTTCTTCAGAACACAACAA 230
|| || || || || || || || || || || || || || || || || || || || ||
QY 163 atcctgtgaagagtgcctcaggaaatgcttcctgctgtggtgagtaagaacagctgt 222
|| || || || || || || || || || || || || || || || || || || || ||
Ddb 231 AACCTGTGAAGAGTGCCTGAAGAAGCTCTCTGCTCTTTGGTGCACACATCAAGGCTTG 290
|| || || || || || || || || || || || || || || || || || || || ||
233 tatggactaccagtaggaaaaatttgcgccctgctttctctctgtaaattgagttccgc 282

Db	291	TCTGGACTACCCAGTTACAAGCGCTTGCACCGGCTTCCTTTGAATTAAGTCTGCG	350
Qy	283	tgcctgggagcgtatgctgggtgaacttcaggcccttgatcataccatcgctgctggg	342
Db	351	ACGCTGGGAGTTGTGTGGGTAACTTTTGAGCGCTGATCATCACCATGCGTAGTCGG	410
Qy	343	gggctgtgcttccctgggcatacactgtgtgctgctactactctgccgcccgaagaag	402
Db	411	GGAAACCCCTCCTCGGCGATTGCCATCTGCCTGCTGCTCTGCTGAGGAGAAGAGGAG	470
Qy	403	ccggaagccagacaagacgcgatgagcggcccatgagagcagggagagagagagtccg	462
Db	471	CCGGAGCGGACAGAGGATGAGGAGNAGGCCATGCGTGAGCCGGAGGAGCGCGATACG	530
Qy	463	gcaggaggaaaggaggcggaatatgaatcaagacatgatgaatacaggaaaaatacgg	522
Db	531	GCAGGAGGAACGGAGACGAGATGAAGACAAGACATGATGAATCAGAAAAAATAATGG	590
Qy	523	tctgtttaagaacaaaacccttatgagaagtcttaa	559
Db	591	CCTGTTTAAGAAGAAAACCCTATGCTAGATTGAA	627
RESULT	2		
LOCUS	AC015892	Mus musculus chromosome 10 clone RP21-386A2 map 10, LOW-PASS	17-NOV-1999
DEFINITION	SEQUENCE SAMPLING.		
ACCESSION	AC015892		
VERSION	AC015892.1 GI:6446776		
KEYWORDS	HTG; HTGS_PHASE0; NULL.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
JOURNAL	Unpublished		
REFERENCE	Mus musculus chromosome 10, clone RP21-386A2		
AUTHORS	2 (bases 1 to 106687)		
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckgater,A., Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,Holland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J.J., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-Nov-1999) Whitehead Institute/MIT Center for Genomics Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html ----- Project Information Center project name: L764 Clone name: 386_A.2		

* NOTE: This record contains 124 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2000, 08:22:03 ; Search time 3493.2 Seconds
(without alignments)
867.931 Million cell updates/sec

Title: US-09-276-268-1
Perfect score: 803
Sequence: 1 gtctgaatggagcatcag.....aaacgacattctgtgagatt 803

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST.*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	560.6	69.8	588	69	AW107669
2	420.8	52.4	457	36	AA646983
					AA646983 ve45a06.i

QY 777 attattaaacgacattctgtgagatt 803
Db 48 ATTATTAAACGACATTCTGTGAGAT 22

RESULT 2
AA646983 457 bp mRNA EST 28-OCT-1997
LOCUS ve45a06.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
DEFINITION IMAGE:821074 5', mRNA sequence.
ACCESSION AA646983
VERSION AA646983.1 GI:2573412
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 457)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407346.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnlni.gov) for further information.
MGI:489354
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 431.
FEATURES
source
1. 457
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:821074"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7N3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 77 a 135 c 130 g 115 t
ORIGIN

Query Match 52.4%; Score 420.8; DB 36; Length 457;
Best Local Similarity 98.9%; Pred. No. 7.5e-113;
Matches 455; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY 79 ggtctggcggtgcagagatggaattctctgcttgggaatcgcggcggttgatctg 138
Db 1 GGTTCTGGCGTGCAGAGATGGACTTCTCGTTCGTTGGATCTCGCGCGGTGATCTG 60

QY 139 tcgggaaactccctgaccagcttccaaaagtccaaggcaggttggcccttcggactctc 198
Db 61 TCGGGAACACTCCCAGACAGCTTCCAAAGFTCAAGGCGAGTTTGG-CCITCGACACTC 119

QY 199 gacctccgcagaaactctctcacggccctccctccagagggtgtgtgtccgagacgctctg 258
Db 120 GACCTCCGAGAAACTCTCTCACGCGCCCTCCCTCAGAGGGTGTGTCCGAGACGCTCTG 179

QY 259 aggggtgtgagacattctacctcagcagaaaccttatgactgtgtggtggaagga 318
Db 180 AGGGGTCTGCAGA-CATCTACCTCAGCAGAACCTTATGACTGCTCTGGGTGGAAGGA 238

QY 319 tgggggccctgcagacgacacttcaagactgttgcgactgttccatggtcaattgcaac 378
Db 239 TGGGGGCCCTGCAGCAGCACACTTCAAGACTGTTGCGGAGTTGTCATGTCCTTGCAC 298

QY 379 ctctctccaagatcgctccgtgtgtgtgagctgcgcgaagcctgcctcagggctgtaag 438
Db 299 CTCTCTCCAAGATGCTCCGCTGGTGGAGCTGCCCGAA-GCCTGCTCAGGCTGTAAAG 357

QY 439 tgggaacagtgagacactgtctcttctacactgtgtctcctgcctcctgcacgtcctcacc 498
Db 358 TGGGAACAGGTGACACTGCTCTCTTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417

QY 499 ctctggtggcctgtactgtctgtcttctcctcacttcaaga 538
Db 418 CTCTGCTGGCCTGTACTGTCTGCTCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 457

RESULT 3
AI505719 381 bp mRNA EST 11-MAR-1999
LOCUS vk51e08.x1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
DEFINITION IMAGE:958214 3', mRNA sequence.
ACCESSION AI505719
VERSION AI505719.1 GI:4403570
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 381)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On May 7, 1998 this sequence version replaced gi:3121530.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnlni.gov) for further information.
MGI:547006
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
High quality sequence stop: 366.
FEATURES
Location/Qualifiers
1..381
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:958214"
/clone_lib="Stratagene mouse Tcell 937311"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: blood; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: